



UNIVERSIDAD
COMPLUTENSE
MADRID

Proyecto de Innovación y Mejora de la Calidad Docente

Convocatoria 2015

Proyecto nº 222

**Simulaciones sanitarias para el estudio de enfermedades
infecciosas endémicas, emergentes y reemergentes en el
ganado porcino**

Responsable: José Manuel Sánchez-Vizcaino Rodríguez

Facultad de Veterinaria

Departamento de Sanidad Animal

1. Objetivos propuestos en la presentación del proyecto

El proyecto propuesto persigue la elaboración de un simulador sanitario interactivo para que los estudiantes de veterinaria puedan practicar sobre la detección temprana de enfermedades infecciosas porcinas y diagnosticar de forma rápida y eficaz las principales enfermedades, tanto endémicas como emergentes o reemergentes. Además de su detección temprana (tema clave en sanidad) deberán establecer los tratamientos más adecuados para su control y erradicación.

Para ello, el proyecto propone utilizar diferentes escenarios virtuales, donde se presenta un problema sanitario con la descripción y visualización del cuadro clínico (fotos y vídeos) y la conversación con el ganadero (elección de las frases más importantes de la conversación). A partir de este hecho, los alumnos tendrán que ir tomando decisiones sobre diferentes posibles líneas de actuación: visitar a los animales, hacer o no necropsias, decir qué muestras son importantes para enviar a laboratorio, qué solicitar al laboratorio, qué diagnóstico diferencial proponen, qué tratamiento inicial proponen, interpretar los resultados del laboratorio, tomar las oportunas acciones en la granja, informar al ganadero, comunicar o no a las autoridades regionales o nacionales, etc. En definitiva, repasar cientos de casos de diferentes enfermedades: endémicas, reemergentes o emergentes, que puedan presentarse en nuestro país y tomar decisiones de acción sobre cada una de ellas.

El planteamiento inicial del proyecto fue basar el simulador sanitario en árboles de decisiones e inferencias bayesianas, así como ejecutar el mismo como una aplicación web, a fin de facilitar su integración con el campus virtual de la UCM. Se planteó, así mismo, que dicho simulador integrara, además, un lenguaje informático específico para el modelado de los árboles de decisión, a fin de facilitar su puesta a punto y mantenimiento por parte de los docentes en veterinaria, así como también su potencial aplicación a otros escenarios educativos (medicina, biología, geología ... y, en general, todos aquellos escenarios en los que tenga sentido plantear simulaciones educativas interactivas de procesos de diagnóstico y de clasificación).

En definitiva, se trata de un proyecto interdisciplinar para la innovación educativa y la mejora de la docencia de las enfermedades infecciosas porcinas. Se ha elegido comenzar con esta especie por ser la de mayor importancia en nuestro país. Actualmente el sector porcino español es el segundo más importante de Europa y el cuarto del Mundo, con una exportación actual de más del 45% de su producción y donde la sanidad es clave para continuar con su crecimiento y expansión.

De esta forma, el proyecto se articula en base a los tres siguientes objetivos:

1. Establecer una herramienta educativa novedosa mediante la creación de un portal web integrado en el campus virtual, de fácil acceso para los alumnos y de potencial aplicación en diferentes disciplinas.
2. Acercar al alumno de una manera práctica e interactiva a la detección temprana de las diferentes enfermedades infecciosas del ganado porcino, haciéndole participe en la toma de decisiones y en el establecimiento de los

tratamientos más adecuados para el control y erradicación de estas enfermedades.

3. Aplicación práctica de modelos epidemiológicos compartimentales y métodos bayesianos para la simulación, predicción y toma de decisiones.

2. Objetivos alcanzados

Nos hemos centrado en la realización del enfoque basado en árboles de decisión y modelos bayesianos para la construcción del simulador interactivo, en la evaluación de dicho enfoque, y en el refinamiento del mismo en base a los resultados de evaluación obtenidos.

De esta forma, en lo que se refiere a la realización del enfoque:

1. Hemos diseñado un primer modelo de ejecución de las simulaciones, basándonos en un modelo de diagnóstico bayesiano clásico. Los requisitos del modelo diseñado son partir de: (i) las *probabilidades a priori* de ocurrencia de las distintas enfermedades, (ii) un árbol de decisión que facilite el diagnóstico de dichas enfermedades, y (iii) probabilidades de explicación de síntomas por enfermedades para cada uno de los nodos hoja de dicho árbol. Utilizando todo este conocimiento, que debe ser proporcionado explícitamente por el experto, el simulador es capaz de mantener una conversación con el alumno, y, en base a las probabilidades condicionales de los nodos hoja alcanzados, y a las probabilidades a priori proporcionadas, el sistema es capaz de calcular las probabilidades a posteriori de las distintas enfermedades.
2. Hemos ideado una forma de codificar todo el conocimiento requerido con Excel, una herramienta ofimática ampliamente utilizada por el equipo de Veterinaria implicado en el proyecto, así como hemos comprobado la factibilidad práctica de procesar el conocimiento codificado para generar la simulación.
3. Hemos construido un prototipo de simulador utilizando tecnologías web del lado del cliente. El simulador toma como entrada una codificación apropiada del conocimiento descrito en la hoja de cálculo Excel, genera un diálogo adecuado con el alumno, calcula probabilidades y arroja diagnósticos posibles ordenados en orden decreciente de probabilidad.

En lo que se refiere a la evaluación del enfoque:

4. La evaluación se ha llevado a cabo con los miembros del equipo de Veterinaria. Los resultados han evidenciado, no obstante, que el enfoque explícito planteado, en el que debe codificarse explícitamente el árbol de decisión, así como las distintas probabilidades requeridas, es excesivamente complejo para ser aplicado en la práctica. Como resultado se ha visto, por tanto, la necesidad de adoptar un enfoque más útil para los expertos veterinarios.

En lo que se refiere al refinamiento del enfoque:

5. Hemos diseñado un nuevo modelo que, por una parte, conserve las características básicas de las simulaciones interactivas (diálogo con el alumno

dirigido por un árbol de decisión, generación de posibles diagnósticos y ponderación de los mismos en base a su *probabilidad*), pero, que por otra parte, no requiera la elicitación de un conocimiento tan explícito por parte de los expertos en Veterinaria (que encuentran bastante tedioso la codificación explícita del árbol de decisión, y que también encuentran difícil, y muchas veces inviable, proporcionar explícitamente las probabilidades). El nuevo modelo es un modelo *basado en casos*. Cada caso consiste en una *ficha* que caracteriza una enfermedad en términos de: (i) la edad del individuo afectado, (ii) el sistema afectado, (iii) los signos clínicos afectados, y (iv) las lesiones observadas. De esta forma, el único conocimiento que deben proporcionar los expertos son estas fichas, conocimiento que se ha comprobado que los expertos pueden proporcionar de manera sencilla. El sistema analiza, entonces, las fichas disponibles para generar automáticamente un árbol de decisión, en términos de los valores presentados por las distintas fichas, que permite guiar el diálogo con el alumno. Como resultado de dicho diálogo se construye un caso observado, que se compara con las distintas fichas estimando, para cada una de ellas, un grado de ajuste. Dichos grados de ajuste se utilizan, entonces, para proporcionar y ordenar los distintos diagnósticos.

6. Hemos ideado una codificación en Excel compacta para las distintas fichas de las enfermedades.
7. Hemos iniciado las modificaciones en el prototipo desarrollado para que funcione sobre el nuevo modelo ideado.

3. Metodología empleada en el proyecto

Tal y como planteamos en la propuesta inicial del proyecto, hemos seguido una metodología de desarrollo del sistema iterativa e incremental, que involucra las siguientes fases:

- Construcción del sistema de simulación. Dicho sistema aporta la infraestructura informática necesaria para la creación de las simulaciones interactivas, incluyendo un formato que permite codificar a los expertos en Veterinaria todo el conocimiento necesario. Estas tareas son realizadas por los miembros de la Facultad de Informática del equipo.
- Construcción del simulador sanitario. Esta fase implica el uso del sistema de simulación para construir el simulador específico objeto de este proyecto (simulador sanitario para el estudio de las enfermedades infecciosas endémicas, emergentes y reemergentes en el ganado porcino). Este trabajo es realizado por los miembros de la Facultad de Veterinaria del equipo.
- Evaluación del simulador sanitario. El simulador construido se evalúa tanto con docentes como con alumnos de Veterinaria. Dicha evaluación se lleva a cabo en tres ejes diferentes: (i) evaluación de la adecuación del sistema de simulación para la creación de las simulaciones educativas (evaluación con docentes), (ii) evaluación de satisfacción de uso del simulador con los alumnos de Veterinaria (evaluación cualitativa orientada a comprobar en qué grado el simulador resulta atractivo, adecuado y usable para los alumnos), (iii) evaluación de eficacia educativa con alumnos de Veterinaria (evaluación

cuantitativa del grado de mejora e impacto del simulador en el proceso educativo).

Como ya se ha indicado, el desarrollo es iterativo e incremental, procediendo a través de diferentes iteraciones y utilizando los resultados de evaluación en cada iteración para mejorar tanto el sistema de simulación como el simulador sanitario en sí. Más concretamente, durante el período de ejecución del proyecto hemos realizado una primera iteración, e iniciado una segunda. En la primera iteración:

- Hemos construido el sistema, basándonos en la codificación explícita de los árboles de decisión y en la provisión explícita de las probabilidades.
- Hemos utilizado el sistema para proporcionar el conocimiento necesario en los términos indicados (árbol de decisión y probabilidades).
- Hemos obtenido resultados de evaluación en el primero de los ejes que, como ya hemos comentado, han evidenciado la complejidad práctica de abordar dicho proceso de adquisición de conocimiento de manera tan explícita.

En la segunda iteración:

- Hemos refinado el modelo para simplificar el proceso de adquisición de conocimiento (el conocimiento requerido se ha reducido a fichas de enfermedades con un formato muy bien definido).
- Hemos iniciado el refinamiento del sistema para adaptarlo al nuevo modelo planteado.

4. Recursos humanos

El equipo de innovación integra a dos miembros de la Facultad de Veterinaria (José Manuel Sánchez-Vizcaíno y Joaquín Goyache) y a dos miembros de la Facultad de Informática (José Luis Sierra y Antonio Sarasa), composición necesaria para el desarrollo de un proyecto interdisciplinar como el propuesto. Efectivamente:

- El desarrollo del sistema de simulación (las herramientas necesarias para crear los simuladores y los motores necesarios para ejecutarlos) implica experiencia en desarrollo de software, experiencia aportada por los Profesores Sierra y Sarasa de la Facultad de Informática. Ambos profesores son miembros del Grupo de Investigación en Ingeniería de Lenguajes Software y Aplicaciones, grupo especializado en desarrollo de software dirigido por lenguajes, y, por tanto, en el tipo de sistema que se propone en este proyecto (de hecho, el Profesor Sierra es el director de dicho grupo).
- El desarrollo del simulador, así como su evaluación pedagógica, implica una experiencia profunda en el dominio del diagnóstico y tratamiento de enfermedades infecciosas en especies animales (en concreto, en la cabaña porcina). A este respecto, el equipo integra a dos miembros del grupo de investigación VISAVET (Centro de Vigilancia Sanitaria Veterinaria), grupo de investigación de referencia en el campo tanto a nivel nacional como internacional: el Profesor Sánchez-Vizcaíno (director del grupo VIVASET, y líder del proyecto), y el Profesor Goyache.

5. Desarrollo de las actividades

Las actividades desarrolladas durante el período de ejecución se han orientado a diseñar el modelo, a proporcionar soporte informático al mismo, a crear las simulaciones utilizando el soporte informático, y a evaluar los productos obtenidos. Como ya se ha indicado anteriormente, dichas actividades se han desarrollado de manera iterativa. De esta forma, durante la primera iteración, y en lo que se refiere a la actividad de diseño del modelo, se ha llevado a cabo la formulación del modelo en los términos explícitos indicados anteriormente (árbol de decisión y probabilidades). Más concretamente, las entradas al sistema consisten en:

- El árbol de decisión. Dicho árbol integra dos tipos de nodos. *Categorías y valores*. Las categorías se corresponden con los distintos aspectos que intervienen en el proceso de diagnóstico (por ejemplo, los anteriormente citados ejes de *edad del individuo*, *sistema afectado*, *signo clínico* y *lesión observada*). Los valores se corresponden con los valores concretos observados para cada una de estas categorías (p.e., *lechazo*, *lechón*, etc. en lo que se refiere a las edades del individuo, o *respiratorio*, *digestivo*, etc. en lo que se refiere al sistema afectado). La raíz del árbol es una categoría. Los hijos de las categorías son valores, que, bien son nodos hoja, bien tienen como hija otra categoría con la que continuar el diálogo.
- Las probabilidades a priori de cada enfermedad contemplada. Dichas probabilidades son probabilidades absolutas de ocurrencia de las enfermedades, y pueden estimarse en base a las frecuencias de ocurrencias de las enfermedades en el área geográfica de estudio (región, país etc.)
- La probabilidad, para cada nodo hoja del árbol h , y cada enfermedad, e , de que la enfermedad e explique los síntomas observados en el camino que lleva a h (es decir, que la enfermedad e explique que un individuo de una determinada edad presente un determinado signo clínico y/o lesión en un sistema dado).

El simulador utiliza entonces esta información para:

- Llevar a cabo una conversación con el alumno: En cada paso, el sistema presenta al alumno los posibles valores para la categoría actual (inicialmente la categoría actual es la raíz del árbol de decisión). El alumno puede seleccionar uno o varios de estos valores, y el sistema continúa el diálogo con las categorías hijas de los valores seleccionados.
- Proporcionar los diagnósticos más probables. Una vez finalizado el diálogo, se habrán alcanzado un conjunto posible de nodos hoja, cada uno de ellos explicados por una serie de enfermedades, con sus probabilidades asociadas. Todas estas probabilidades se combinan utilizando análisis bayesiano, bajo la suposición de independencia condicional entre explicaciones de caminos de diagnóstico por enfermedades, para inferir una probabilidad a posteriori para cada enfermedad.

En lo que se refiere al soporte informático proporcionado, durante la primera iteración:

- Se ha ideado el formato Excel para la codificación del conocimiento de diagnóstico al que se ha hecho alusión anteriormente. Más concretamente, dicho formato fija que:
 - Debe proporcionarse una tabla en la que, para cada enfermedad, se indica su probabilidad a priori.
 - Para cada camino en el árbol de decisión (codificado en el formato *Categoria.Valor.Categoria.Valor ...*) debe proporcionarse una tabla encabezada por dicho camino, y que, para cada enfermedad, proporcione la probabilidad de que dicha enfermedad explique los síntomas que figuran en el camino.
- Se ha construido un traductor de estos datos Excel a estructuras JSON procesables por el motor del simulador, utilizando PHP.
- Se ha construido un prototipo de dicho motor utilizando JavaScript, CSS3 y HTML5.

En lo que se refiere a la creación de las simulaciones utilizando el soporte informático desarrollado, se ha creado un árbol de diagnóstico inicial, se han fijado probabilidades iniciales de acuerdo con los requisitos del modelo, y todo ello se ha codificado como una hoja Excel siguiendo el formato establecido.

En lo que se refiera a la evaluación, se han identificado las principales fortalezas y debilidades del enfoque propuesto:

- Por una parte, se ha comprobado que el comportamiento interactivo de las simulaciones, consistente en un diálogo inicial en el que se recaba información sobre los síntomas del caso, seguido de un diagnóstico ordenado por probabilidad, es el adecuado de cara a su uso por parte de los alumnos.
- Por otra parte, los principales problemas se han detectado, como ya se ha indicado, en el enfoque de adquisición de conocimiento:
 - Por una parte, la descripción exhaustiva del árbol de diagnóstico puede resultar poco práctica (o incluso impracticable) conforme aumenta el número de enfermedades y de síntomas. Así mismo, dicha descripción puede ser difícil de depurar, mantener y refinar.
 - Por otra parte, mientras que los expertos en Veterinaria no han encontrado difícil proporcionar las probabilidades condicionales asociadas con las hojas del árbol, sí han encontrado más difícil proporcionar las probabilidades a priori de las enfermedades, de manera absoluta. Así mismo, también han encontrado bastante tedioso proporcionar todas las probabilidades requeridas, siendo el proceso bastante poco escalable y bastante propenso a errores.

Como resultado de esta primera iteración, hemos comenzado una segunda iteración orientada a resolver las debilidades detectadas en el enfoque de adquisición de conocimiento durante la evaluación. Para ello, hemos emprendido de nuevo la actividad de diseño del modelo, con el fin de relajar los requisitos de conocimiento para los expertos veterinarios. De esta forma:

- Tal y como ya hemos indicado anteriormente, hemos prescindido totalmente de codificaciones explícitas de árboles, o de establecimiento explícito de

probabilidades. En su lugar, lo único que exigimos ahora es recolectar *fichas clínicas* que caractericen enfermedades en términos de los ejes de edad, sistema afectado, signos clínicos y lesiones ya citados anteriormente. Hemos ratificado, así mismo, que este tipo de información sí puede ser proporcionada de manera cómoda y natural por los expertos en veterinaria.

- Hemos formulado un nuevo modelo de ejecución que, por una parte, preserve el modelo de interacción ya implementado en la iteración previa, y que, por otra, sea capaz de tratar con fichas de enfermedades, en lugar de con codificaciones explícitas de árboles y probabilidades. Básicamente, dicho modelo toma como entradas:
 - El árbol de diagnóstico (este árbol se genera automáticamente a partir de las fichas, como describiremos seguidamente).
 - Una descripción de las fichas en términos de: (i) la edad del individuo afectado, (ii) el conjunto de signos clínicos referidos en la ficha, y (iii) el conjunto de lesiones.
 - Una tabla de pesos, que mide la relevancia de cada signo clínico y cada lesión a efectos de discriminar entre enfermedades (de esta forma, aquellos signos y lesiones que ocurren menos frecuentemente entre enfermedades se consideran más relevantes de cara a la discriminación que los que ocurren más frecuentemente, y, por tanto, tendrán asignados un peso mayor)

De esta forma, el proceso de diálogo con el alumno es idéntico al del modelo previo. No obstante, la fase de estimación de probabilidades se substituye por una medida de similitud con fichas. Para ello:

- Se seleccionan primeramente aquellas fichas que se corresponden con la edad observada (al resto se le asigna una similitud 0)
- Para cada ficha seleccionada, se construye un vector de pesos, con un peso por cada signo o lesión. Aquellos signos y lesiones ausentes se pesan como 0. El peso del resto se determina a partir de la tabla de pesos.
- Se construye el correspondiente vector de pesos para los signos y lesiones observados.
- Se miden similitudes calculando el coseno del vector de pesos observado con cada vector de pesos de cada ficha. Dichas similitudes juegan el papel de las probabilidades en el anterior enfoque bayesiano.
- Hemos ideado una forma compacta y sencilla de usar para proporcionar las fichas de enfermedades en Excel. Básicamente, se utiliza una única tabla, cuyas filas se corresponden con edades, signos clínicos y lesiones de sistemas afectados, y cuyas columnas se encabezan con enfermedades. Por tanto, cada columna es una ficha. Con **X** se marcan las edades, signos y lesiones de la ficha.
- Estamos modificando el generador de simulaciones PHP para tratar con el nuevo modelo. Para ello, el generador:
 - Mide la relevancia de cada signo y lesión utilizando su frecuencia inversa en las fichas (la medida es análoga a la utilizada en los modelos clásicos de recuperación de la información).

- Construye automáticamente el árbol de diagnóstico partiendo de un árbol vacío y filtrando sucesivamente fichas en el mismo. Así mismo, utiliza los pesos de los atributos para ordenar adecuadamente los valores en dicho árbol.
 - Vuelca toda la información en estructuras JSON, que pueden ser procesadas por el motor del simulador.
- Estamos modificando, por último, el motor del simulador para que implemente el nuevo modelo de ejecución basado en fichas.

En Madrid, 30 de Abril de 2016



Prof. Dr. José Manuel Sánchez-Vizcaíno Rodríguez